

Structural basis of preinitiation complex assembly on human Pol II promoters

Francis T.F.Tsai and Paul B.Sigler¹

Department of Molecular Biophysics and Biochemistry and the Howard Hughes Medical Institute, Yale University, New Haven, CT 06511, USA

¹Corresponding author
e-mail: sigler@csb.yale.edu

Transcription initiation requires the assembly of a preinitiation complex (PIC), which is nucleated through binding of the TATA-box binding protein (TBP) to the promoter. Biochemical studies have shown, however, that TBP recognizes the TATA-box in both orientations and, therefore, cannot account for the directionality of PIC assembly. Transcription factor IIB (TFIIB) is essential for transcription initiation from RNA polymerase II promoters. Recent functional studies have identified a specific 7 bp TFIIB recognition element (BRE) immediately upstream of the TATA-box. We present here the 2.65 Å resolution crystal structure of a human TFIIBc–TBPc complex bound to an idealized and extended adenovirus major late promoter. This structure now reveals that human TFIIBc binds to the promoter asymmetrically through base-specific contacts in the major groove upstream and in the minor groove downstream of the TATA-box. Binding of TFIIBc is, therefore, synergistic with TBPc requiring the distortion of the TATA-box. Thus, the newly described TFIIBc–DNA interface is likely to be a key determinant for the unidirectional assembly of a functional PIC.

Keywords: DNA recognition/TBP/TFIIB/transcription/
X-ray crystallography

Introduction

Transcription initiation of all protein encoding genes in Eukarya requires the formation of a preinitiation complex (PIC), which consists of RNA polymerase II (Pol II) and the basal transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIIF and TFIIH (reviewed in Ranish and Hahn, 1996; Roeder, 1996). TFIID is a multi-protein complex, consisting of the TATA-box binding protein (TBP) at its core, and so-called TBP-associated factors (TAFs), which are essential for regulated transcription *in vitro* (reviewed in Verrijzer and Tjian, 1996; Sauer and Tjian, 1997), but are not required for basal transcription (Hoffmann *et al.*, 1990; Martinez *et al.*, 1994). It is generally accepted that binding of TBP to the TATA-box nucleates the formation of the PIC either through a stepwise assembly of other basal factors, or through recruitment of a holoenzyme (reviewed in Koleske and Young, 1995; Orphanides *et al.*, 1996; Pugh, 1996). It remains unclear, however, what determines the orientation of PIC assembly and, thus, how the

direction of transcription initiation is determined (reviewed in Cox *et al.*, 1998; Reinberg *et al.*, 1998; Tsai *et al.*, 1998). Although TAFs have been reported to recognize an initiator element (Nakatani *et al.*, 1990; Martinez *et al.*, 1994; reviewed in Smale *et al.*, 1998), the initiator does not determine the transcriptional directionality in the presence of a TATA element (O'Shea-Greenfield and Smale, 1992; reviewed in Smale *et al.*, 1998). In support of this notion, earlier biochemical studies have shown that the methidiumpropyl-EDTA-Fe(II) footprint of TFIID alone was centered symmetrically on a consensus TATA sequence (Sawadogo and Roeder, 1985), indicating that contacts between TAFs and the DNA do not confer transcriptional polarity. Conversely, similar studies using a functional PIC clearly revealed an asymmetric footprint on the promoter (Van Dyke *et al.*, 1988), suggesting that other basal transcription factors determine the orientation of PIC assembly and, therefore, the direction of transcription initiation.

The structures of several eukaryal basal transcription factors, alone and in complex with DNA, have been determined by X-ray crystallography and NMR (Nikolov *et al.*, 1992, 1995, 1996; Kim *et al.*, 1993a,b; Bagby *et al.*, 1995; Geiger *et al.*, 1996; Juo *et al.*, 1996; Tan *et al.*, 1996). Together, these structures have shown at the molecular level how the core domain of TBP (TBPc) recognizes the TATA-box, and through its severe distortion of the DNA nucleates the formation of the PIC. Biochemical studies have shown, however, that TBPc recognizes the TATA-box in both orientations (reviewed in Cox *et al.*, 1998; Reinberg *et al.*, 1998; Tsai *et al.*, 1998). This is supported by the numerous crystal structures of a TBPc–TATA-box complex that reveal pseudo-2-fold rotational symmetry of the TBP–DNA interface (reviewed in Tsai *et al.*, 1998). Several models have been proposed by which polarity of TBP binding is achieved. This includes an asymmetry of the electrostatic charge potential coupled to the deformability of the TATA element (Kim *et al.*, 1993b; Kosa *et al.*, 1997), and a slight asymmetry in the stereochemistry of the TBPc–TATA-box interface (Kim *et al.*, 1993a; Juo *et al.*, 1996). However, promoter affinity cleavage studies revealed only a modest 60:40 preference for the 'correct' orientation when yeast TBPc (yTBPc) binds to either adenovirus major late (MLP) or yeast CYC1 promoter (Cox *et al.*, 1997). Moreover, the crystal structure of a highly homologous archaeal TF(II)Bc–TBP complex on a short promoter fragment (containing only 2 bp upstream of the TATA-box) shows the TF(II)Bc–TBP complex to be bound in an inverted orientation relative to its eukaryal counterparts (Kosa *et al.*, 1997).

Mutagenic selection, *in vitro* transcription and *in vitro* binding assays, in addition to cross-linking experiments using human factors, indicated that binding of TFIIB to the MLP may be highly asymmetrical (Lagrange *et al.*,

Table I. Helical parameters of the MLP

DNA	Helical twist (°)	Roll (°)	Tilt (°)	Rise (Å)	Propeller twist (°)	Buckle (°)
<i>Gua-Cyt</i> (−41)	31.11	4.50	−0.01	3.46	7.91	4.94
<i>Gua-Cyt</i> (−40)	28.46	5.43	1.93	3.57	13.61	4.50
<i>Gua-Cyt</i> (−39)	37.16	3.03	−4.39	3.80	14.64	1.30
<i>Gua-Cyt</i> (−38)	27.94	7.90	−2.44	3.08	4.16	−10.41
<i>Gua-Cyt</i> (−37)	33.62	11.62	2.34	3.33	0.98	−5.15
<i>Gua-Cyt</i> (−36)	32.20	−2.30	−3.74	3.28	4.10	0.60
<i>Cyt-Gua</i> (−35)	37.22	2.36	6.90	3.43	−12.66	4.79
<i>Gua-Cyt</i> (−34)	32.15	6.46	−0.67	3.15	−5.34	2.98
<i>Cyt-Gua</i> (−33)	37.80	−0.09	−5.60	3.83	−7.22	9.52
<i>Cyt-Gua</i> (−32)	34.48	2.94	4.24	3.19	−4.30	−8.66
Thy-Ade (−31)	17.65	42.25	0.56	5.98	15.80	−2.06
Ade-Thy (−30)	17.33	25.67	2.64	4.06	18.34	−39.34
Thy-Ade (−29)	20.11	9.96	3.54	2.94	−3.27	−30.96
Ade-Thy (−28)	6.32	28.68	−0.21	3.49	6.56	−3.35
Ade-Thy (−27)	18.91	22.20	−0.69	4.03	9.58	17.20
Ade-Thy (−26)	19.57	24.12	−0.52	3.72	1.34	21.23
Ade-Thy (−25)	14.32	36.82	−2.83	5.81	−6.90	30.13
Gua-Cyt (−24)	30.79	6.85	0.00	3.46	13.86	4.75
<i>Gua-Cyt</i> (−23)	30.52	5.29	2.09	3.55	7.42	7.41
<i>Gua-Cyt</i> (−22)	27.29	4.38	−2.13	3.51	11.42	2.86
<i>Gua-Cyt</i> (−21)	39.73	0.90	−1.92	3.76	12.79	0.35
<i>Gua-Cyt</i> (−20)	29.90	5.34	−4.59	2.96	1.85	−10.99
<i>Gua-Cyt</i> (−19)	35.04	10.21	3.69	3.36	5.33	−1.16
<i>Gua-Cyt</i> (−18)	30.95	0.23	−4.04	3.29	7.06	1.88
<i>Cyt-Gua</i> (−17)	35.38	1.36	6.56	3.40	−7.69	1.90
<i>Gua-Cyt</i> (−16)	–	–	–	–	−5.46	3.24

Helical parameters of the idealized MLP in a representative complex of the crystal structure as determined with the program CURVES 5.1 (Lavery and Sklenar, 1988) using a local helix axis to calculate the helical twist, roll, tilt and rise. The DNA sequence is shown from upstream to downstream relative to the transcription start site as illustrated schematically in Figure 3. The TATA-box is highlighted in bold, and bases of adjacent ternary complexes are shown in italics. The black bars indicate the boundaries of the DNA in each ternary complex.

1996, 1998). These studies suggested that the core domain of human TFIIB (hTFIIBc) recognizes a 7-bp-long, so-called TFIIB recognition element (BRE) immediately upstream of the TATA-box with the consensus sequence: 5'-G/C-G/C-G/A-C-G-C-C-3' (Lagrange *et al.*, 1998; Reinberg *et al.*, 1998). Since TFIIB is an essential factor in the assembly of a functional PIC, and forms a stereospecific complex with TBP (Tansey and Herr, 1997), it would follow that a specific TFIIB–BRE interaction would contribute strongly to unique directionality in the assembly of the PIC and, hence, to the polarity of transcription.

This orientation would be consistent with the preferred alignment of TBP on the TATA-element, as defined by complementary altered specificity mutations in the TBP–DNA interface (Strubin and Struhl, 1992). Functional studies similar to those of Lagrange *et al.* (1998) have also been carried out in Archaea (Qureshi and Jackson, 1998) where the basal transcription complex is highly homologous to that of Pol II. These studies identified a BRE of slightly different consensus sequence, which appears to form a specific interaction with the TFIIB homolog, TFB (Qureshi and Jackson, 1998; Tsai *et al.*,

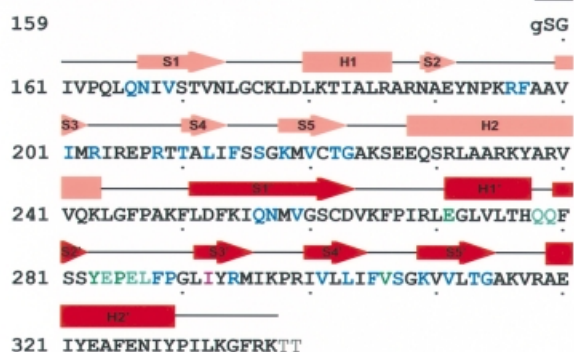
A Sequence of hTFIIBc**B Sequence of hTBPc****C Sequence of MLP**

Fig. 1. Components of the hTFIIBc-hTBPc-MLP complex. Sequences of hTFIIBc (A), hTBPc (B) and idealized MLP (C) used for crystallization. The light and dark hues of the bars (α -helices) and arrows (β -sheets) represent the N- and C-terminal domains, respectively. The same colors and hues are used throughout in all figures. Sequence numbers refer to the full-length protein with the initiation Met being the first residue. A black dot beneath the sequence marks every 10 amino acids. Amino acid residues seen in the crystal structure are shown in upper-case letters and bold face, residues that are disordered in upper-case letters only, and residues that are part of a leader sequence in lower-case letters. Residues in hTFIIBc highlighted in red contact hTBPc (A), and those in hTBPc highlighted in green interact with hTFIIBc (B). Residues that contact the DNA, as depicted schematically in Figure 3, are highlighted in blue and those that make both protein and DNA contacts are in purple (A and B). An asterisk in (A) indicates a residue in hTFIIBc that makes the same contact with the DNA in the previously determined structure (Nikolov *et al.*, 1995). A gray box in (C) highlights the TATA-box and arrows indicate the sequence of the BRE.

1998). It is noteworthy that the archaeal BRE is located in precisely the same position on the promoter as its eukaryal counterpart and, indeed, has been shown to determine the direction of transcription in Archaea (Bell *et al.*, 1999).

The previous crystal structures of a eukaryal ternary TFIIBc-TBP2-TATA-box complex and its archaeal

counterpart did not reveal any specific contacts between TF(II)Bc and the major or minor groove edges of bases within the promoter (Nikolov *et al.*, 1995; Kosa *et al.*, 1997). In view of the current structural results presented here, and recent studies of the BRE in Eukarya (Lagrange *et al.*, 1998) and Archaea (Qureshi and Jackson, 1998), this is presumably because of the inadequate length and non-ideal sequence of the promoter segments flanking the TATA-box in the oligonucleotides used for crystallization.

To understand what determines the polarity of PIC assembly and, thus, the direction of transcription on human and perhaps other TATA-box-containing Pol II promoters in metazoans, we have crystallized a complex consisting of hTBPc and hTFIIBc bound to an idealized MLP containing a version of the BRE identified by selection binding experiments (Lagrange *et al.*, 1998; Reinberg *et al.*, 1998). This structure differs from the previously determined structure of the *Arabidopsis thaliana* TBP2-hTFIIBc-DNA complex (Nikolov *et al.*, 1995) in that both proteins are of human origin, and most importantly, in the length and sequence of the promoter fragment. Here, the DNA extends 6.5 bp upstream and 3.5 bp downstream of the TATA-box. In fact, the promoter fragment extends even further because in this crystal lattice the DNA packs end-to-end in perfect helical register through a non-contiguous 5' single-base-pair overlap (see Table I for helical parameters). The crystal structure reveals that, in addition to extensive contacts with the phosphoribose backbone, hTFIIBc makes specific base contacts in both the major groove of the BRE and the minor groove immediately downstream of the TATA-box. The upstream interactions are mediated through a helix-turn-helix (HTH) motif, similar to that found in bacterial HTH DNA-binding proteins as pointed out by Lagrange *et al.* (1998) and Reinberg *et al.* (1998). The downstream interactions involve a 'recognition-loop' linking helices BH2 and BH3 in the N-terminal cyclin-like repeat of hTFIIBc, and reveal a previously unreported base-specific interaction with the promoter.

Results**Structure of a human TFIIBc-TBPc-MLP complex**

Figure 1A-C illustrates the sequences of hTFIIBc, hTBPc and the idealized MLP used for crystallization of the human ternary complex. There are five ternary complexes in the asymmetric unit linked, as noted above, through a 1 bp overlap that produces a covalently discontinuous but conformationally contiguous DNA helix. As each complex has a different crystal packing environment, consistent stereochemical features are inherent to the complex and independent of the crystal lattice. Many of the features of this complex are similar to those in the previously described complex of *A.thaliana* TBP2 and hTFIIBc bound to a shorter MLP (Nikolov *et al.*, 1995). Human TBPc forms an interface with the TATA-box similar to those seen in all TBPc-DNA complexes solved to date (Figure 3; Kim *et al.*, 1993a,b; Geiger *et al.*, 1996; Juo *et al.*, 1996; Nikolov *et al.*, 1996; Tan *et al.*, 1996). The deformation of the DNA is similar in all five of the current representations in the asymmetric unit when superimposed pairwise [the root mean square deviation (r.m.s.d.) for all atoms of the TATA-box is 0.14 ± 0.03 Å], and deviates from the

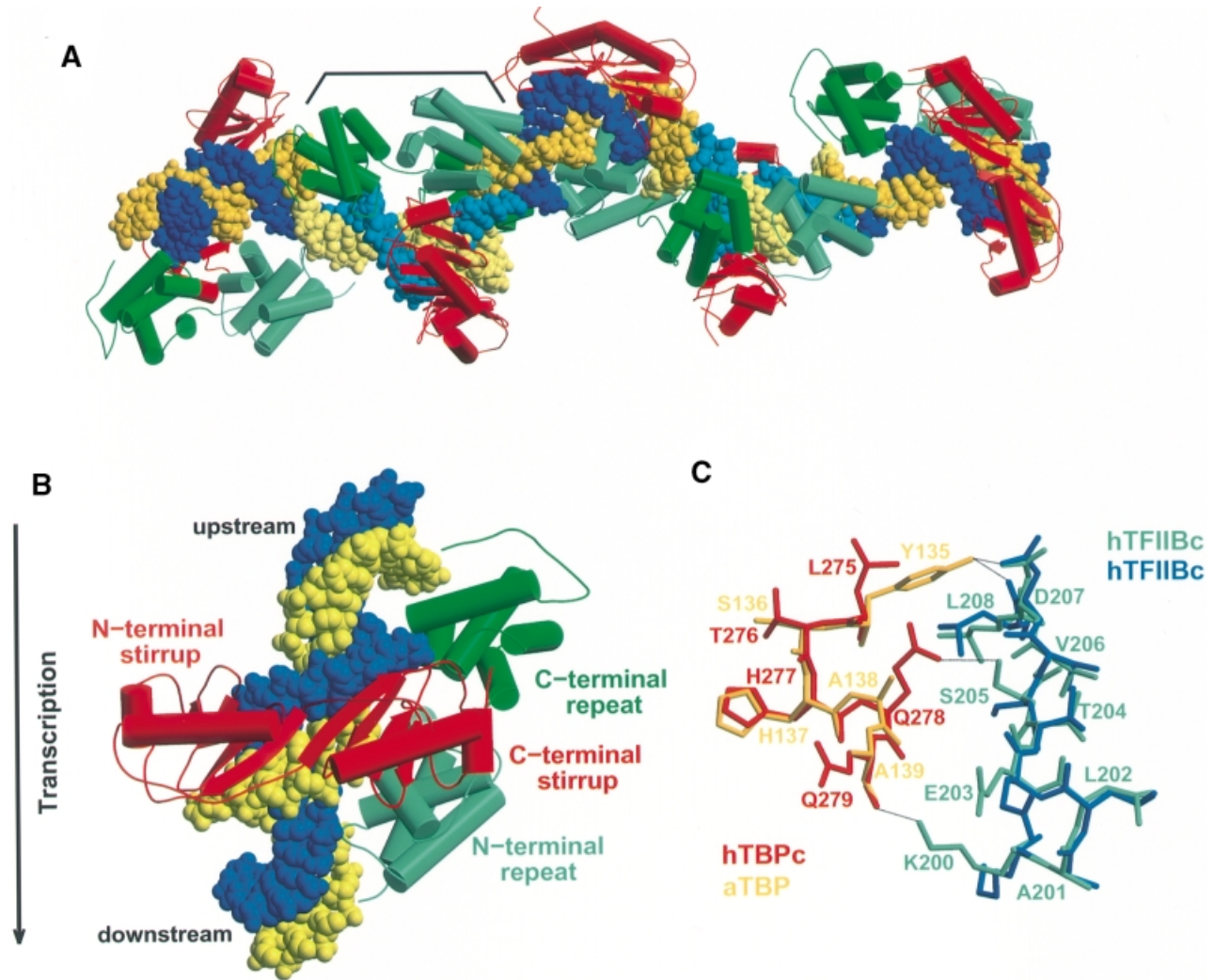


Fig. 2. Architecture of the human ternary TFIIBc-TBPc-MLP complex. (A) Ribbons and space-filling representation of one asymmetric unit. The five ternary complexes are distinguished by alternating dark and light yellow (coding strand) and dark and light blue (non-coding strand). These color designations are maintained throughout. The crystal is formed of ‘infinite’ strands of essentially B-DNA angulated in the 8 bp TATA-box through contact with hTBPc (see Table I for helical parameters). The non-contiguous helical union occurs at a 5’ single base pair overlapping sticky end. A representative ternary complex is bracketed. (B) An isolated ternary complex as viewed down the pseudo-2-fold axis of the hTBPc-TATA-box interaction. The surfaces of the C-terminal repeat of hTFIIBc that are in contact with the BRE are shown in dark green, and those that are in contact with the minor groove in light green. The arrow indicates the direction of transcription. (C) Species-specific differences in the TBP-TFIIB interface. Residues of hTBPc are colored in red, and those of *A.thaliana* (aTBP2) in yellow. Residues of hTFIIBc are shown in light green, and those of the previously determined structure in blue. Hydrogen bonds are indicated by a dotted line. Figures were generated with MOLSCRIPT/RASTER3D (Kraulis, 1991; Merritt and Bacon, 1997).

deformation imposed on the TATA-box by *A.thaliana* TBP2 in the previously solved ternary complex (Nikolov *et al.*, 1995) by an r.m.s.d. of only 0.60 ± 0.02 Å. However, in the structure reported here (see below), hTFIIBc forms a much larger and more intimate interface with the DNA flanking the TATA-box as a result of the difference in length and sequence of the DNA used for crystallization. Moreover, due to the fortuitous, non-covalent, in-register extensions of the TATA-box flanking segments (Table I), our structure provides direct and complete visualization of the interactions between hTFIIBc and the BRE reported by Lagrange *et al.* (1998), as well as a previously unreported base-specific interaction with the minor groove downstream of the TATA-box (Figure 3).

Figure 2A shows the components of an asymmetric

unit as seen in the crystal structure of the human TFIIBc-TBPc-DNA complex. The figure depicts the orientation of all five complexes, which are stacked head to tail, and are held in continuous helical register by the G(-38) and C(-20’) base pair produced by the 5’ overhang. Upon binding to DNA, the five copies of the ternary complex bury a total of $25\,900$ Å² of solvent-accessible area, or ~ 5200 Å² per complex, nearly 1000 Å² more than the 4200 Å² in the previously determined crystal structure (Nikolov *et al.*, 1995). The five representations of the hTFIIBc-hTBPc complex reported here show an r.m.s.d. of 0.41 ± 0.28 Å when superimposed pairwise through their C_α atoms. The C_α atoms of the ternary complex superimpose on their counterparts in the previously reported structure with an r.m.s.d. of only 0.84 ± 0.19 Å,

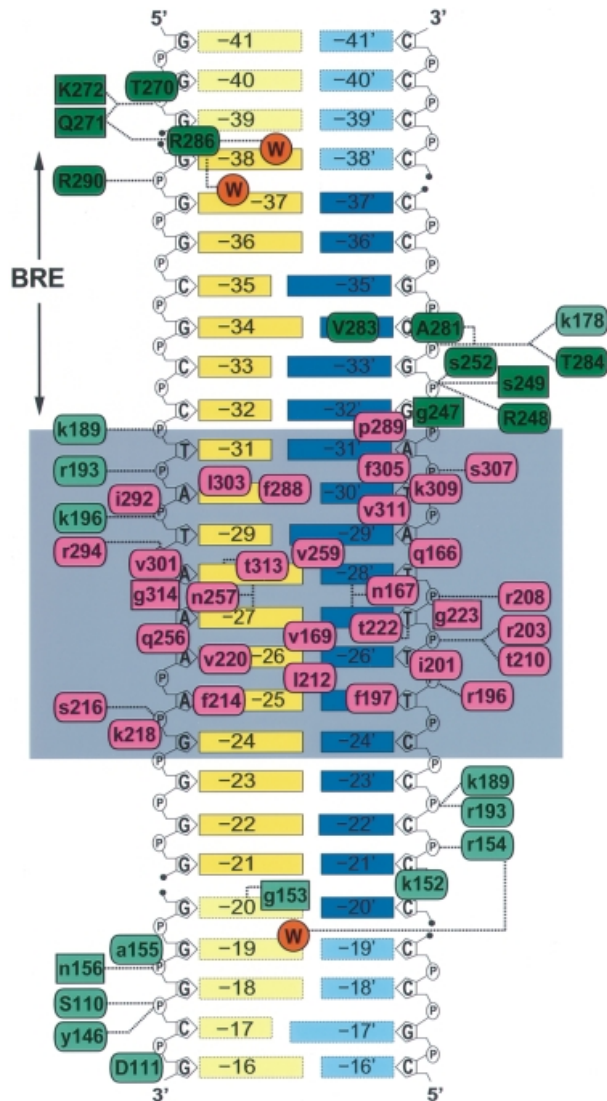


Fig. 3. Protein–DNA contacts that specify the orientation of the hTFIIBc–hTBPc–MLP complex. Complete schematic illustrating all protein–DNA interactions in the ternary complex. Arrows indicate the location of the BRE. An oval indicates an interaction between the promoter and the protein side chain, and a square an interaction with the protein main chain. Arg154 and Ala281 of hTFIIBc contact MLP through both main and side chain. Amino acid residues that are in contact with the major groove are shown in upper-case letters, and those in contact with the minor groove in lower-case letters. The DNA is numbered with respect to the transcription start site. The non-coding strand is designated by a prime throughout this paper. Hydrogen bonds are represented by dotted lines. The color designation of the protein and DNA components is explained in Figures 1 and 2A. Water molecules that mediate contacts with the DNA bases are shown as orange circles. A gray box highlights the TATA-box.

indicating that the protein components and their relative orientations are essentially the same in the two structure determinations. This result is somewhat surprising in view of (i) the significant differences in the hTFIIBc–DNA interface; and (ii) the species difference in TBP that caused a minor mismatch in the protein–protein interface between the TBP2 from *A.thaliana* and hTFIIBc (Figure 2C).

Interaction between hTFIIBc and MLP

As a result of the perfect helical register of the DNA segments that join adjacent ternary complexes (Table I),

hTFIIBc can make stereochemically appropriate interactions with the promoter as far as 9 bp upstream and 8 bp downstream of the TATA-box (Figure 3). This, in turn, leads to a more intimate juxtaposition of hTFIIBc and the DNA throughout the promoter. For example, in this structure, the side chains of Lys189 and, in two of the complexes, Arg193 of hTFIIBc each make hydrogen bonds with the phosphate backbone at the upstream end of the TATA-box [T(–31) and A(–30), respectively], and immediately downstream of it at C(–23') (Figure 4). These bridging, polar interactions between hTFIIBc and the promoter require the severe deformation of the TATA-box, but presumably contribute to the stability of the ternary complex, and thereby reinforce the TBP-induced deformation of the DNA. Thus, while the deformed TATA-box of the current structure and that of Nikolov *et al.* (1996) have the same TBPC interface, and essentially the same DNA backbone conformation as noted above (r.m.s.d. of 0.60 ± 0.02 Å), the DNA segments flanking the TATA-box have very different hTFIIBc interfaces and, hence, very different backbone conformations (r.m.s.d. of 2.94 ± 0.23 Å for the phosphoribose backbone).

Human TFIIBc interacts with the major groove upstream of the TATA-box through an HTH motif in a manner similar to that described for bacterial proteins. This HTH motif (Figure 5A–C) is supported by helix BH3', and consists of helices BH4' and BH5', the so-called 'recognition-helix'. The specific interactions between helix BH5' and the BRE are illustrated in Figure 5A. In our structure, Val283 makes a van der Waals interaction with the C5–C6 edge of C(–34'), 3 bp upstream of the TATA-box, in all five copies of the ternary complex. A G:C base pair at this position, i.e. 3 bp upstream of the consensus TATA-box, is conserved in 32% of all human and 56% of all human viral Pol II promoters that have been deposited in the Eukaryotic Promoter Database (EPD) (Cavin Périer *et al.*, 1998). The importance of this van der Waals interaction is further underscored by mutagenesis data, which indicated that Val283 confers specificity for a G:C base pair at position –34 of the BRE (Lagrange *et al.*, 1998). This G:C base pair is also the strongest determinant in the sequence selection experiment, which led to the identification of the BRE (Lagrange *et al.*, 1998). In addition, the side chain of a highly conserved and mutational sensitive Arg286 (Lagrange *et al.*, 1998) makes water-mediated contacts with the edges of bases of G(–38) in four copies, additionally with G(–37) in two copies, and van der Waals interactions with the phosphoribose backbone of G(–38) in all five copies of the complex. The water-mediated interactions, as opposed to a direct interaction, are most compatible with the 6-keto hydrogen bond acceptor of the preferred guanine base but can also accept a hydrogen bond donated by an adenine. In a 'buttressing' pattern remarkably similar to those employed by bacterial/phage HTH DNA-binding motifs, the long and potentially flexible side chain of Arg286 is stabilized by the side-chain carbonyl oxygen of a conserved glutamine (Gln271) located at the first position of helix BH4'. This glutamine is, in turn, immobilized through a bifurcated hydrogen bond between the main-chain amide and the phosphoribose backbone of G(–39), and through a hydrogen bond between its side-chain amide nitrogen and the hydroxyl group of Tyr259 (Figure 5A). The

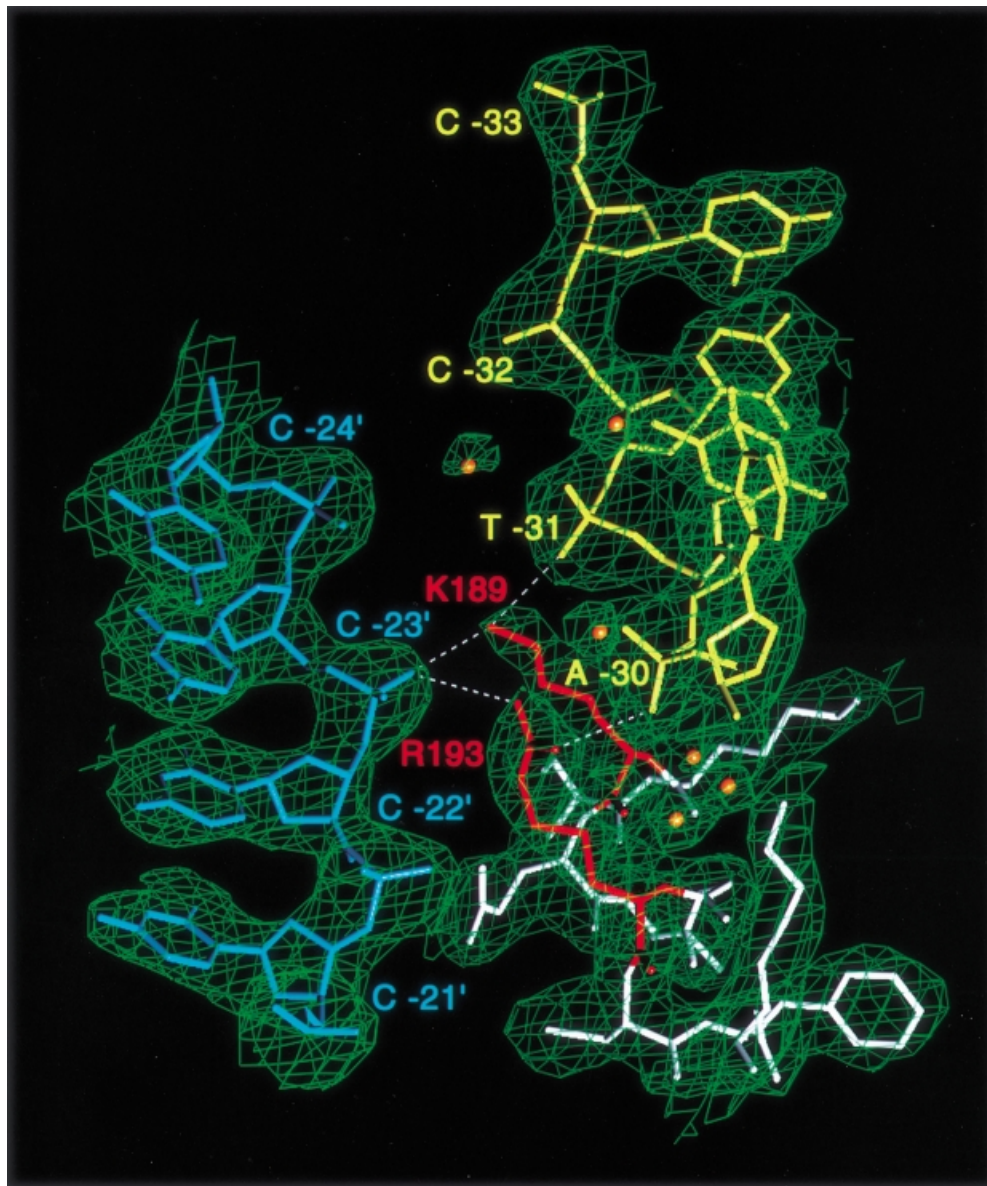


Fig. 4. Interaction between hTFIIBc and the promoter that stabilizes the hTBPc-induced deformation of the TATA-box. hTFIIBc is shown in white, the DNA strands in yellow (coding) and blue (non-coding), and water molecules in orange. Lys189 and Arg193, which make hydrogen bonds (dotted line) with the DNA backbone, are colored in red and labeled accordingly. The sigma weighted $2F_o - F_c$ map (green) was calculated with phases from the refined coordinates of the human TFIIBc-TBPc-DNA complex and is contoured at a 1.1 sigma level. The figure was generated using TURBO-FRODO (Roussel and Cambillau, 1989).

interface resembles that of HTH-containing bacterial/phage regulatory proteins in two other ways. First, the interaction of the HTH motif with the major groove edges of bases is bracketed by a network of hydrogen bonds and van der Waals contacts between the polar side chains or backbone amides of conserved residues in and around the HTH, and the phosphoribose backbone on both sides of the major groove (Figures 3 and 5A). Secondly, a small positive roll in the recognition surface between the G:C base pairs at positions -36 and -37 is also noticeable (Table I). A sequence comparison reveals that several residues of this interface, namely Arg248, Ser249, Gln271, Thr284, Arg286, Gln287 and Arg290, are also conserved or are very similar among metazoan and archaeal TF(II)Bs, suggesting that the interactions of this HTH motif have

evolved from a common ancestor. This notion is supported by the recently determined crystal structure of an archaeal TBP-TF(II)B-DNA complex (Tsai *et al.*, 1998; Littlefield *et al.*, 1999) that shows a similar interaction between the HTH and the BRE. Interestingly, in yeast and plants Val283 is replaced by a glycine, suggesting that TFIIB may recognize the BRE differently in these organisms.

The contacts between hTFIIBc and the minor groove of the MLP downstream of the TATA-box are mediated through a loop (residues 152-156) that links helices BH2 and BH3 in the first cyclin repeat of hTFIIBc (Figure 6A and B). This loop is anchored to the minor groove through a base-specific interaction between the carbonyl oxygen of Gly153 and the 2-amino group of G(-20) base in all five copies, and via a water-mediated contact between the

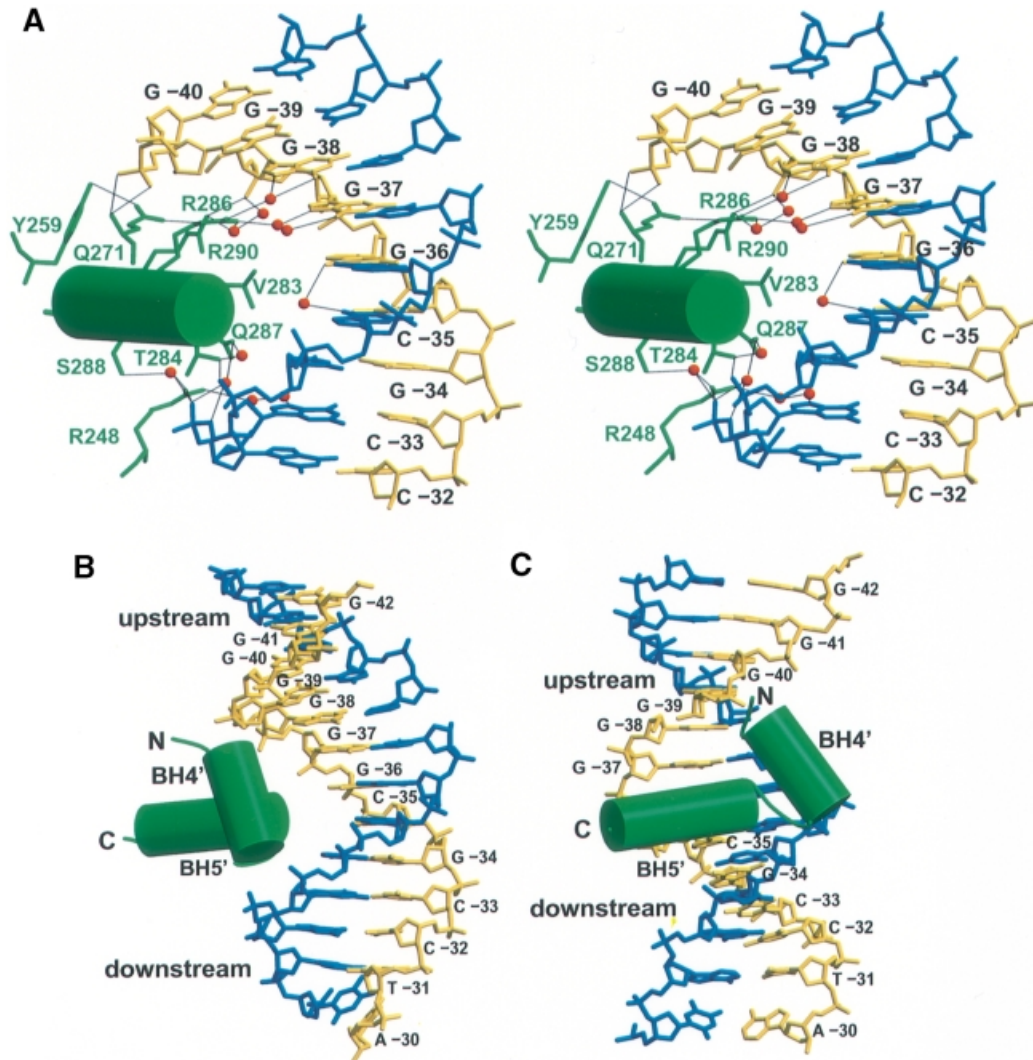


Fig. 5. (A) Stereo view depicting the molecular interactions between the ‘recognition-helix’ BH5’ (green) of the HTH and the BRE. Amino acid residues (green) that contact the major groove are labeled accordingly. For clarity, only the bases of the coding strand (yellow) are labeled. Water molecules are shown as orange spheres. A black line indicates hydrogen bonds. (B and C) HTH motif consisting of helices BH4’ and BH5’ of hTFIIBc. The view in (C) is similar to (B) rotated by $\sim 90^\circ$ around the DNA helical axis. Label and color designations are as described in (A) and Figures 1 and 2A. Figures were generated with MOLSCRIPT/RASTER3D (Kraulis, 1991; Merritt and Bacon, 1997).

carbonyl oxygen of Arg154 and the 3-amino group of G(-19) in four copies of the complex. Moreover, the side chains of Lys152, Arg154, Ala155 and Asn156 reinforce the interaction between this recognition-loop and the minor groove by forming both van der Waals and polar contacts with the DNA backbone on both sides of the groove (Figures 3 and 6A).

Interestingly, because of the continuous helical packing of the promoter fragments in the crystal structure described here, the major groove upstream of the TATA-box (extended by 3 bp upstream into the neighboring promoter) is occupied by the C-terminal cyclin-like repeat of hTFIIBc, and the minor groove is occupied by the N-terminal cyclin-like repeat of an hTFIIBc molecule bound to the adjacent promoter. It is totally fortuitous that the deformation of the TATA-box and the helical register of the promoter fragments enable the proteins from adjacent complexes to have access to the DNA of their neighbor on opposite faces of the helix. This also suggests that,

in vivo, this minor groove surface of the promoter is available for interaction with other components of the basal transcriptional machinery, such as full-length hTFIIA, which has been shown to cross-link to the MLP upstream of the TATA-box in a human TBP-TFIIA-TFIIIB-DNA quaternary complex (Lagrange *et al.*, 1996).

Comparison of hTFIIBc alone and in complex with hTBPc-DNA

The overall structure of hTFIIBc in the crystalline ternary complex differs from that seen in the NMR structure of hTFIIBc alone (Bagby *et al.*, 1995). Although each cyclin-like repeat of free hTFIIBc shares the same fold with the corresponding repeat of hTFIIBc in the ternary complex, the two domains are arranged differently in the two structures (Hayashi *et al.*, 1998). Superimposing free hTFIIBc and hTFIIBc in the crystalline complex through their N-terminal cyclin-like repeats reveals that the respective HTH DNA-binding motifs located in the C-terminal

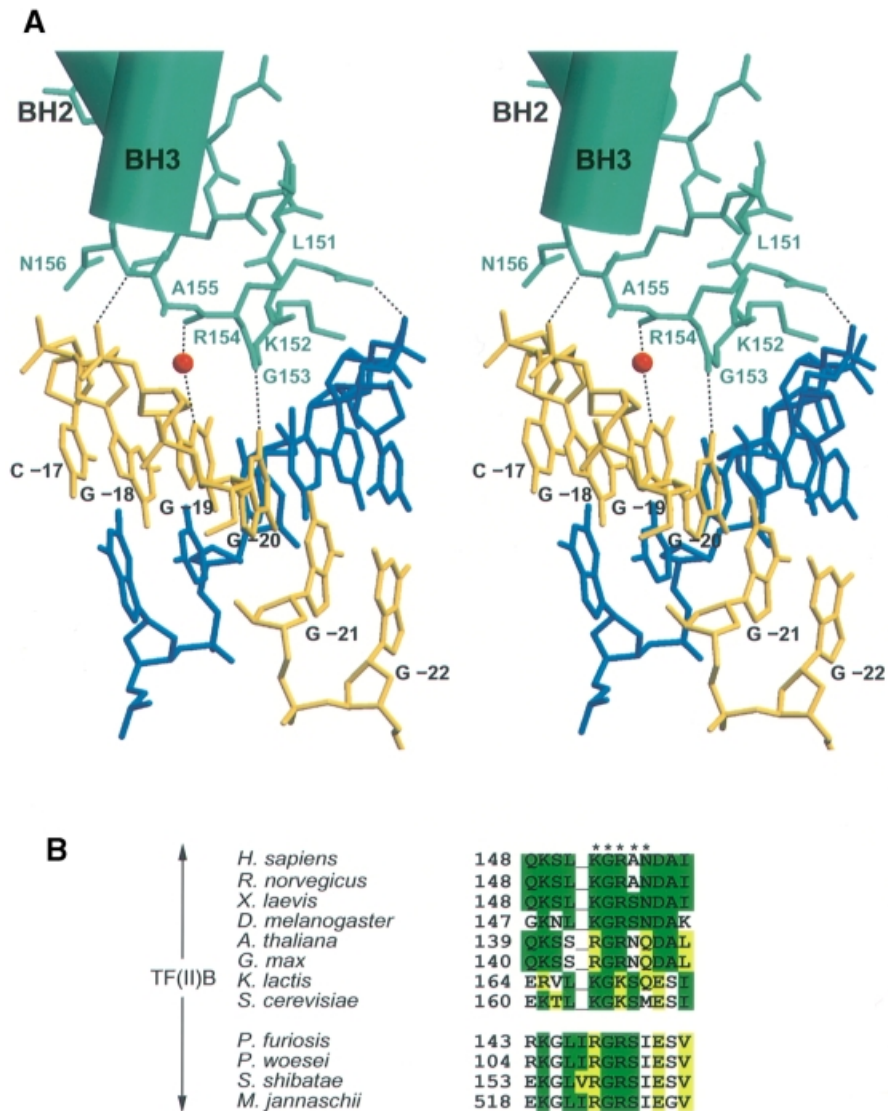


Fig. 6. (A) Stereo view depicting the molecular interactions between the minor groove recognition-loop (residues 152–156 of hTFIIBc) and the MLP immediately downstream of the TATA-box. The color designation is the same as in Figures 1 and 2A. For clarity, only the bases of the coding strand (yellow) are labeled. The figure was generated with MOLSCRIPT/RASTER3D (Kraulis, 1991; Merritt and Bacon, 1997). (B) Alignment of the minor groove recognition-loop sequences of eukaryal and archaeal TF(II)Bs. Green highlighting indicates conserved, and yellow highly similar residues. An asterisk indicates a residue in hTFIIBc that contacts the minor groove in the current structure.

repeat point in opposite directions. Hence, in order for free hTFIIBc to form functionally important, specific promoter contacts, the cyclin-like repeats must undergo a large *en bloc* conformational rearrangement (a rotation of $\sim 100^\circ$) upon forming a competent ternary complex. This rearrangement of the two cyclin-like domains orients the two surfaces of TFIIB correctly in order to interact simultaneously with the major groove upstream and the minor groove downstream of the TATA-box.

Discussion

The crystal structure of the human TFIIBc–TBPc–MLP complex presented here describes the stereochemistry of the interface formed by hTFIIBc and both the major groove of the BRE and the minor groove downstream of

the TATA-box. These interactions are consistent with a unique polarity of the ternary complex, and the assembly of the PIC, which would define the direction of transcription in the absence of other basal factors. In order for these contacts to occur in concert, deformation of the TATA-box by hTBPc is essential. The TBP-distorted DNA is further stabilized by Lys189 and possibly Arg193 of hTFIIBc, each of which forms a pair of interactions that span the arched TATA-box (Figure 4). Moreover, the simultaneous binding of the HTH in the C-terminal cyclin-like repeat to the major groove upstream, and the binding of the recognition-loop of the N-terminal cyclin-like repeat to the minor groove downstream, is compatible only with the TBP-induced deformation of the TATA-box, suggesting that binding of hTFIIBc and hTBPc to the promoter is synergistic. Thus, the crystal structure presented here

reveals that hTFIIBc, aided by the deformation of the DNA by hTBPc, 'differentiates' between the major groove upstream and the minor groove immediately downstream of the TATA-box through base-specific contacts and, therefore, provides polarity for the nucleating events in the assembly of a competent PIC on TATA-box-containing promoters.

Structural basis of transcriptional polarity

Recent *in vitro* biochemical studies in human (Lagrange *et al.*, 1996, 1998) and archaeal systems (Qureshi and Jackson, 1998) indicate that the 6–7 bp segment contiguous with the upstream boundary of the TATA-box provides a sequence-specific binding site for TF(II)B and, thus, a basis for asymmetric assembly of the PIC. A search of all metazoan Pol II promoter sequences deposited in the EPD (Cavin Périer *et al.*, 1998), however, reveals that there is great diversity among Pol II promoters, even within a single species. Indeed, some Pol II promoters contain no TATA-box at all and, thus, probably no BRE either. The diversity of promoter sequences, coupled with the wide variety of protein–protein contacts involved in recruitment and modulation of PIC assembly and function, supports the assertion of Cox *et al.* (1997) that the orientation of the PIC is through the accrual of specific interactions between various factors and the promoter, each contributing a fraction of the binding energy that defines the correct orientation of the PIC. These factors include basal transcription factors such as TFIIB, as revealed from this work, but also TAFs and other components of the holoenzyme, which may play a more dominant role on TATA-less Pol II promoters (Martinez *et al.*, 1994; Burke and Kadonaga, 1996; Burke *et al.*, 1998). Nevertheless, the crystal structure presented here reveals that the assembly of the PIC on a TATA-containing metazoan promoter is only compatible with the 'correct' orientation of hTBPc on the promoter. A PIC assembly in the reverse orientation would clearly violate the base-specific contacts observed in the hTFIIBc–DNA interface and, therefore, be energetically unfavorable.

The functional studies on human Pol II promoters by Lagrange *et al.* (1996, 1998) suggest that the BRE, which itself is only modestly conserved, provides a preferential, albeit somewhat variable, binding surface for hTFIIBc. The variability within the sequence of the BRE is consistent with our finding that the hydrogen-bonded interactions between hTFIIBc and the edges of bases of BRE's major groove are through water-mediated contacts. This situation is reminiscent of the relaxed stereospecificity provided by a hydrated contact surface in the interface between the DNA-binding domain of the estrogen receptor and its half-site (Schwabe *et al.*, 1993; Xu *et al.*, 1993). This type of interface permits increased degrees of freedom in forming stable interactions with a weakly conserved target, in a recognition system where other geometric factors contribute to the stability.

Materials and methods

Expression and purification of protein components

The 180-amino-acid-residue C-terminal core domain of human TBP was overexpressed and purified essentially as described in Juo *et al.* (1996), with the addition of a heparin–Sepharose CL-6B (Amersham Pharmacia

Biotech) chromatography step following the initial Ni–NTA resin (Qiagen) column.

The plasmid encoding the C-terminal core domain of human TFIIB (pET21d/hTFIIBc), consisting of amino acid residues 107–316 linked to an N-terminal leader sequence (Met–Gly–His₇–Ser–Gly–Leu–Val–Pro–Arg–Gly–Ser–Arg–Thr), was a gift of S. Juo (Yale University). The protein was overexpressed in *Escherichia coli* BL21 (DE3) (Novagen) in Luria–Bertani medium in the presence of 0.1 mg/ml ampicillin. Cells were grown at 37°C, induced at OD₆₀₀ = 1.0 with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and harvested 4 h after induction. Human TFIIBc was purified via a two-step protocol consisting of Ni–NTA resin (Qiagen) chromatography, followed by an SP-Sepharose fast flow (Amersham Pharmacia Biotech) chromatography step. The N-terminal leader sequence was removed by thrombin cleavage (Haematologic Technologies Inc.) prior to the second chromatographic step, leaving an artifactual four-amino-acid N-terminal sequence, namely Gly–Ser–Arg–Thr, preceding Met107.

The plasmid encoding the 78-amino-acid-residue VP16 acidic activation domain, VP16 AAD (pGEX-3X/VP16 AAD), from the type I herpes simplex virus fused to a glutathione *S*-transferase (GST)-tag was a gift from S. Triezenberg (Michigan State University). The protein was overexpressed in *E. coli* BL21 (DE3) (Novagen) in Terrific-Broth medium in the presence of 0.1 mg/ml ampicillin. Cells were grown at 30°C, induced at OD₆₀₀ = 1.0 with 0.1 mM IPTG, and harvested 4–5 h after induction. The cell pellet was resuspended in 25 mM Tris–HCl pH 7.5, 0.2 mM EDTA and 0.2 M KCl, to which phenylmethylsulfonyl fluoride and benzamidine–HCl were added to final concentrations of 2 mM each. Cells were lysed using a microfluidizer processor (Microfluidics). The cell debris was removed by ultracentrifugation at 75 000 g for 45 min. The fusion protein was purified via a glutathione–Sepharose 4B (Amersham Pharmacia Biotech) affinity chromatography step. Following this, the GST-tag was removed by Factor Xa cleavage (Haematologic Technologies Inc.), and separated from the VP16 AAD by a Q-Sepharose fast flow (Amersham Pharmacia Biotech) chromatography step.

Synthesis, purification and annealing of oligonucleotides used for crystallization

Oligonucleotides used for crystallization were synthesized via the phosphoramidite method (Keck facility, Yale University), and purified by hydrophobic interaction/affinity chromatography (Nensorb Prep, DuPont). After purification, the oligonucleotides were annealed by slow cooling from 95 to 4°C at 0.2°C/min using the MiniCycler (MJ Research Inc.) in 10 mM Tris–HCl pH 8.0, 50 mM KCl and 5 mM MgCl₂, resulting in a final stock concentration of 50 mM double-stranded DNA as estimated by measuring the absorbance at 260 nm (calculated $\epsilon_{260} = 337\,000\text{ M}^{-1}\text{cm}^{-1}$).

Preparation and crystallization of the hTFIIBc–hTBPc–DNA complex

All protein components were concentrated and dialyzed against 40 mM Tris–HCl pH 8.0, 300 mM ammonium acetate, 50 mM KCl, 5 mM MgCl₂, 5 mM dithiothreitol (DTT) and 10% glycerol using a Centricon-10 and -3 (Amicon), respectively. The molarities of the protein solutions were estimated by measuring their absorbance at 280 nm [calculated by the method of Gill and von Hippel (1989); $\epsilon_{280} = 6760\text{ M}^{-1}\text{cm}^{-1}$ for hTFIIBc and $7860\text{ M}^{-1}\text{cm}^{-1}$ for hTBPc] in 6.0 M guanidine–HCl, 20 mM potassium phosphate buffer at pH 6.5. The hTFIIBc–hTBPc–DNA complex was prepared at a stoichiometric ratio of 1:1:1.5 at 1/10 of the final concentration used for crystallization, and incubated on ice overnight. Following this, the ternary complex was concentrated and mixed with a 1.5 M excess of VP16 AAD ($\epsilon_{280} = 2560\text{ M}^{-1}\text{cm}^{-1}$), resulting in a final concentration of ~0.16 mM (10 mg/ml) of the protein–DNA complex.

hTFIIBc–hTBPc–DNA in the presence of VP16 AAD was crystallized by the hanging drop vapor-diffusion method by mixing an equal volume of protein sample with well solution consisting of 4% PEG 8000 (w/v), 50 mM Tris–HCl pH 8.0, 50 mM MgCl₂, 100 mM sodium citrate pH 5.6 and 10% glycerol. Crystals (Parent 1) appeared within 2 days at 4°C, and reached a typical size of 0.6 × 0.1 × 0.05 mm after 4–6 weeks.

The crystals were harvested into a stabilizer solution consisting of 8% PEG 8000, 45 mM Tris–HCl pH 7.5, 150 mM ammonium acetate, 27.5 mM magnesium acetate, 25 mM KCl, 50 mM sodium citrate pH 5.6, 22.5 mM DTT and 10% glycerol, and were transferred in series into stabilizer solutions containing a progressively higher concentration of glycerol (15–35% final). The crystals were suspended in small nylon loops (Hampton Research) at the end of Yale mounting pins, and flash-

Table II. Data collection and refinement statistics

Data and refinement statistics	Parent 1 CHESS F2	Parent 2 BNL X25	Parent 3 CHESS F1
Resolution ^a	50–2.65 Å (2.71–2.65 Å)	35–2.95 Å (3.02–2.95 Å)	30–3.2 Å (3.31–3.20 Å)
Completeness ^a	97.5% (85.4%)	98.1% (95.0%)	97.3% (97.3%)
I/σ^a	16.0 (2.9)	21.5 (3.0)	15.4 (8.3)
$R_{\text{sym}}^{a,b}$	5.7% (34.1%)	4.6% (33.0%)	5.3% (14.0%)
No. of reflections ($F > 2\sigma$)			
$R_{\text{work}}/R_{\text{free}}^c$	90 918/4854	63 168/7107	
$R_{\text{work}}/R_{\text{free}}^c$ ($F > 2\sigma$)	22.9%/26.0%	23.0%/27.6%	
R.m.s.d. from ideality			
bond lengths	0.013 Å	0.015 Å	
bond angles	1.58°	1.72°	
B-factors (r.m.s.d. of bonded atoms: main/side chain)	1.28 Å ² /2.29 Å ²	2.18 Å ² /3.05 Å ²	

^aValues for the highest resolution shell are given in parentheses.

^b $R_{\text{sym}} = (\sum_{\text{hkl}} |I(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} I(\text{hkl}))$, where $\langle I(\text{hkl}) \rangle$ is the mean of the Friedel or symmetry equivalent reflections of $I(\text{hkl})$.

^c5% (Parent 1) and 10% (Parent 2) of all data.

frozen by plunging into liquid propane, which was allowed to solidify in liquid nitrogen for storage.

To confirm whether crystals could be obtained in the absence of VP16 AAD, a complex consisting of hTFIIBc, hTBPc and DNA only was also prepared. Isomorphous crystals (Parent 2 and 3) of similar size and morphology were readily obtained under similar conditions to those described above. However, crystals grown in the presence of VP16 AAD diffracted consistently to higher resolution. The final difference Fourier map did not reveal any electron density that could account for VP16 AAD. Moreover, an SDS–PAGE analysis showed that VP16 AAD was not present in the crystals (data not shown). Hence, crystals obtained in the presence of VP16 AAD are referred to throughout the text as crystals of the ternary complex.

Data collection and structure determination

Data of the ternary complex were collected at 100 K on an Area Detector Systems Corporation Quantum-4 CCD (2K × 2K) detector at CHESS F2 (Parent 1) and F1 (Parent 3), on a nine-element CCD (3K × 3K) at APS ID19 (iodine derivative, data not shown), and on a Brandeis B4 CCD (2K × 2K) detector at BNL X25 (Parent 2 and bromine derivative, data not shown), respectively. The CHESS and BNL data were integrated, scaled, and merged with DENZO and SCALEPACK (Otwinowski, 1993), and the APS data were processed with the HKL 2000 suite (Otwinowski and Minor, 1997). The crystals belong to the monoclinic space group, $P2_1$, with unit cell dimensions $a = 118.45$ Å, $b = 122.30$ Å, $c = 140.22$ Å, $\beta = 113.08^\circ$. The asymmetric unit contains five ternary complexes, totaling a mass of ~277 kDa. The structure was determined initially by molecular replacement with the program AMoRe [Navaza, 1994; Collaborative Computational Project 4 (CCP4), 1994] using the coordinates of the previously determined structure of the TFIIBc ternary complex (1VOL; Nikolov *et al.*, 1995) as a search model, and data collected on a Parent 2 crystal. The rotation function was calculated using all data between 8 and 4 Å, and a radius of integration sphere of 40 Å. The rotation function search gave four solutions with peak heights of 11.8, 10.2, 6.5 and 6.0 sigma over the mean. The translation function search compatible with the four top rotation functions gave an overall correlation function of 42.6 and R -value of 44.5%. Subsequent rigid-body fitting of the four complexes improved the overall correlation coefficient and R -value to 43.2 and 44.0%, respectively. Each solution corresponds to one ternary complex in the asymmetric unit. Calculating a rotation function using data collected on a Parent 3 crystal identified a fifth complex. The addition of the fifth complex in the translation function search (correlation coefficient 46.4 and R -value 43.2%), followed by subsequent rigid-body fitting of all five complexes, improved the overall correlation coefficient and R -value to 46.7 and 42.7%, respectively.

The molecular replacement solution was confirmed by difference Patterson and difference Fourier analyses that compared data collected from a Parent 1 crystal and similar crystals containing a halogenated oligonucleotide, in which the two thymidines in the coding strand of the

TATA-box element were substituted with either 5-iodo-dU or 5-bromo-dU.

Structure refinement

Prior to refinement, 5% of all data (Parent 1) were excluded to calculate the R_{free} -value for cross-validation analysis (Brünger, 1992). Each of the five complexes were first refined as rigid bodies in CNS (Brünger *et al.*, 1998), with isotropic temperature factor, and no bulk solvent corrections against all data between 8 and 6 Å. In the subsequent cycles of refinement, the data were slowly extended to higher resolution up to 2.65 Å. With the inclusion of higher resolution data, the model was split into smaller rigid-body units. Once the positions of the five complexes were defined, the model was refined in CNS by several cycles of simulated annealing and positional refinement using non-crystallography symmetry (NCS) constraints, torsion angle molecular dynamics (Rice and Brünger, 1994), and a maximum likelihood target (Adams *et al.*, 1997). The five improper NCS operators were obtained by electron-density averaging techniques using the programs IMP and DM (Cowtan, 1994) of the RAVE (Kleywegt and Jones, 1994) and CCP4 program suites (CCP4, 1994), respectively. The refinement was interspersed by rebuilding the model manually into $2F_o - F_c$ and $F_o - F_c$ difference Fourier maps using the program TURBO-FRODO (Roussel and Cambillau, 1989). The course of refinement was monitored by following the decrease in the R_{free} -value. In the final cycles of refinement, the NCS constraints were replaced by restraints, and the structure was refined by positional refinement using torsion angle molecular dynamics, overall anisotropic temperature factor and bulk solvent corrections. The refinement statistics of the final model are shown in Table II. Continuous electron density was observed for all components of the ternary complex, except for the first three amino acid residues of hTFIIBc and the last two residues of hTBPc. The final model also contained a total of 329 water molecules, of which 172 have common positions in at least two of the five representations of the ternary complex in the asymmetric unit as determined using the program WATNCS (CCP4, 1994). No density for the VP16 AAD was observed in difference Fourier maps. PROCHECK (Laskowski *et al.*, 1993) revealed that 88.2% of residues are in the most favored regions, 11.8% in additional allowed regions, and no residues, except glycines, are in disallowed regions of the Ramachandran plot.

Coordinate deposition

The atomic coordinates and structure-factor amplitudes of the refined hTFIIBc–hTBPc–MLP complex have been deposited with the Protein Data Bank under accession number 1C9B.

Acknowledgements

We are grateful to S.Juo for the pET21d/hTBPc and pET21d/hTFIIBc vectors, and to S.Triezenberg for the pGEX-3X/VP16 AAD plasmid. We thank members of the Sigler laboratory for help with data collection

and useful discussion, the staff of MacCHESS (Drs S.Ealick and D.Thiel), NSLS X25 (Drs L.Berman, H.Lewis and R.Sweet) and APS-SBC 19ID (Dr A.Joachimiak) for aid in obtaining data, and Dr P.Adams for help with CNS. F.T.F.T. is a recipient of a Wellcome Trust International Prize Travelling Research Fellowship (049086/Z/96/Z/JMW/LEC). This work was supported in part by a grant (GM15225) from the National Institutes of Health.

References

- Adams,P.D., Pannu,N.S., Read,R. and Brünger,A.T. (1997) Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.
- Bagby,S., Kim,S., Maldonado,E., Tong,K.I., Reinberg,D. and Ikura,M. (1995) Solution structure of the C-terminal core domain of human TFIIB: similarity to cyclin A and interaction with TATA-binding protein. *Cell*, **82**, 857–867.
- Bell,S.D., Kosa,P.L., Sigler,P.B. and Jackson,S.P. (1999) Orientation of the transcription preinitiation complex in Archaea. *Proc. Natl Acad. Sci. USA*, **96**, 13662–13667.
- Brünger,A.T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.
- Brünger,A.T. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
- Burke,T.W. and Kadonaga,J.T. (1996) *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.*, **10**, 711–724.
- Burke,T.W., Willy,P.J., Kutach,A.K., Butler,J.E.F. and Kadonaga,J.T. (1998) The DPE, a downstream core promoter element that is functionally analogous to the TATA box. *Cold Spring Harb. Symp. Quant. Biol.*, **63**, 75–82.
- Cavin Périer,R., Junier,T. and Bucher,P. (1998) The eukaryotic promoter database EPD. *Nucleic Acids Res.*, **26**, 353–357.
- Collaborative Computational Project 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
- Cowtan,K.D. (1994) DM program. *Joint CCP4 ESF–EACBM Newslett.*, **31**, 34–38.
- Cox,J.M., Hayward,M.M., Sanchez,J.F., Gegnas,L.D., van der Zee,S., Dennis,J.H., Sigler,P.B. and Schepartz,A. (1997) Bidirectional binding of the TATA box binding protein to the TATA-box. *Proc. Natl Acad. Sci. USA*, **94**, 13475–13480.
- Cox,J.M., Kays,A.R., Sanchez,J.F. and Schepartz,A. (1998) Preinitiation complex assembly: potentially a bumpy path. *Curr. Opin. Chem. Biol.*, **2**, 11–17.
- Geiger,J.H., Hahn,S. and Sigler,P.B. (1996) Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science*, **272**, 830–836.
- Gill,S.C. and von Hippel,P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.*, **182**, 319–326.
- Hayashi,F., Ishima,R., Liu,D., Tong,K.I., Kim,S., Reinberg,D., Bagby,S. and Ikura,M. (1998) Human general transcription factor TFIIB: conformational variability and interaction with VP16 activation domain. *Biochemistry*, **37**, 7941–7951.
- Hoffmann,A., Sinn,E., Yamamoto,T., Wang,J., Roy,A., Horikoshi,M. and Roeder,R.G. (1990) Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human TATA factor (TFIID). *Nature*, **346**, 387–390.
- Juo,Z.S., Chiu,T.K., Leiberman,P.M., Baikov,I., Berk,A.J. and Dickerson,R.E. (1996) How proteins recognize the TATA box. *J. Mol. Biol.*, **261**, 239–254.
- Kim,J.L., Nikolov,D.B. and Burley,S.K. (1993a) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**, 520–527.
- Kim,Y., Geiger,J.H., Hahn,S. and Sigler,P.B. (1993b) Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512–520.
- Kleywegt,G.J. and Jones,T.A. (1994) Halloween...masks and bones. In Bailey,S., Hubbard,R. and Waller,D. (eds), *From First Map to Final Model*. SERC Daresbury Laboratory, Warrington, UK, pp. 59–66.
- Koleske,A.J. and Young,R.A. (1995) The RNA polymerase II holoenzyme and its implication for gene regulation. *Trends Biochem. Sci.*, **20**, 113–116.
- Kosa,P.F., Ghosh,G., DeDecker,B.S. and Sigler,P.B. (1997) The 2.1 Å crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. *Proc. Natl Acad. Sci. USA*, **94**, 6042–6047.
- Kraulis,P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Lagrange,T., Kim,T.-K., Orphanides,G., Ebright,Y.W., Ebright,R.H. and Reinberg,D. (1996) Organization of the TBP–TFIIA–TFIIB–DNA quaternary complex defined by site-specific protein–DNA photocrosslinking. *Proc. Natl Acad. Sci. USA*, **93**, 10620–10625.
- Lagrange,T., Kapanidis,A.N., Tang,H., Reinberg,D. and Ebright,R.H. (1998) New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor TFIIB. *Genes Dev.*, **12**, 34–44.
- Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
- Lavery,R. and Sklenar,H.J. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.*, **6**, 63–91.
- Littlefield,O., Korkhin,Y. and Sigler,P.B. (1999) The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc. Natl Acad. Sci. USA*, **96**, 13668–13673.
- Martinez,E., Chiang,C.-M., Ge,H. and Roeder,R.G. (1994) TATA-binding protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. *EMBO J.*, **13**, 3115–3126.
- Merritt,E.A. and Bacon,D.J. (1997) Raster3D: photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.
- Nakatani,Y., Horikoshi,M., Brenner,M., Yamamoto,T., Besnard,F., Roeder,R.G. and Freese,E. (1990) A downstream initiation element required for efficient TATA box binding and *in vitro* function of TFIID. *Nature*, **348**, 86–88.
- Navaza,J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A*, **50**, 157–163.
- Nikolov,D.B., Hu,S.-H., Lin,J., Gasch,A., Hoffmann,A., Horikoshi,M., Chua,N.-H., Roeder,R.G. and Burley,S.K. (1992) Crystal structure of TFIID TATA-box binding protein. *Nature*, **360**, 40–46.
- Nikolov,D.B., Chen,H., Halay,E.D., Usheva,A.A., Hisatake,K., Lee,D.K., Roeder,R.G. and Burley,S.K. (1995) Crystal structure of a TFIIB–TBP–TATA element ternary complex. *Nature*, **377**, 119–128.
- Nikolov,D.B., Chen,H., Halay,E.D., Hoffmann,A., Roeder,R.G. and Burley,S.K. (1996) Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl Acad. Sci. USA*, **93**, 4862–4867.
- Orphanides,G., Lagrange,T. and Reinberg,D. (1996) The general transcription factors of RNA polymerase II. *Genes Dev.*, **10**, 2657–2683.
- O’Shea-Greenfield,A. and Smale,S.T. (1992) Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. *J. Biol. Chem.*, **267**, 1391–1402.
- Otwinowski,Z. (1993) Oscillation data reduction program. In Sawyer,L., Isaacs,N. and Bailey,S. (eds), *Data Collection and Processing*. SERC Daresbury Laboratory, Warrington, UK, pp. 56–62.
- Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–325.
- Pugh,B.F. (1996) Mechanisms of transcription complex assembly. *Curr. Opin. Cell Biol.*, **8**, 303–311.
- Qureshi,S.A. and Jackson,S.P. (1998) Sequence-specific DNA binding by the *S.shibatae* TFIIB homolog, TFB and its effect on promoter strength. *Mol. Cell*, **1**, 389–400.
- Ranish,J.A. and Hahn,S. (1996) Transcription: basal factors and activation. *Curr. Opin. Genet. Dev.*, **6**, 151–158.
- Reinberg,D. *et al.* (1998) The RNA polymerase II general transcription factors: past, present and future. *Cold Spring Harb. Symp. Quant. Biol.*, **63**, 83–103.
- Rice,L.M. and Brünger,A.T. (1994) Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins*, **19**, 277–290.
- Roeder,R.G. (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.*, **21**, 327–335.
- Roussel,A. and Cambillau,C. (1989) TURBO-FRODO. In *Silicon Graphics Geometry Partner Directory*. Silicon Graphics, Mountain View, CA, pp. 77–78.
- Sauer,F. and Tjian,R. (1997) Mechanisms of transcriptional activation: differences and similarities between yeast, *Drosophila* and man. *Curr. Opin. Genet. Dev.*, **7**, 176–181.
- Sawadogo,M. and Roeder,R.G. (1985) Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Nature*, **43**, 165–175.

- Schwabe,J.W.R., Chapman,L., Finch,J.T. and Rhodes,D. (1993) The crystal structure of the complex between the oestrogen receptor DNA-binding domain and DNA at 2.4 Å: how receptors discriminate between their response elements. *Cell*, **75**, 567–578.
- Smale,S.T., Jain,A., Kaufmann,J., Emami,K.H., Lo,K. and Garraway,I.P. (1998) The initiator element: a paradigm for core promoter heterogeneity within metazoan protein-coding genes. *Cold Spring Harb. Symp. Quant. Biol.*, **63**, 21–31.
- Strubin,M. and Struhl,K. (1992) Yeast and human TFIID with altered DNA-binding specificity for TATA elements. *Cell*, **68**, 721–730.
- Tan,S., Hunziker,Y., Sargent,D.F. and Richmond,T.J. (1996) Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature*, **381**, 127–134.
- Tansey,W.P. and Herr,W. (1997) Selective use of TBP and TFIIB revealed by a TATA–TBP–TFIIB array with altered specificity. *Science*, **275**, 829–831.
- Tsai,F.T.F., Littlefield,O., Kosa,P.F., Cox,J.M., Schepartz,A. and Sigler,P.B. (1998) Polarity of transcription on pol II and archaeal promoters: where is the ‘one-way sign’ and how is it read? *Cold Spring Harb. Symp. Quant. Biol.*, **63**, 53–61.
- Van Dyke,M.W., Roeder,R.G. and Sawadogo,M. (1988) Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. *Science*, **241**, 1335–1338.
- Verrijzer,C.P. and Tjian,R. (1996) TAFs mediate transcriptional activation and promoter selectivity. *Trends Biochem. Sci.*, **21**, 338–342.
- Xu,W., Alroy,I., Freedman,L.F. and Sigler,P.B. (1993) Stereochemistry of specific steroid receptor–DNA interactions. *Cold Spring Harb. Symp. Quant. Biol.*, **58**, 133–139.

Received October 18, 1999; revised and accepted November 4, 1999